

- Persael due tomorrow Tuesday 9.1
 - 1.3 and 2.1 and 2.2
- Webwork will be ~~be~~ made available soon (hopefully Tuesday) due Sunday
- Keep trying Gradescope & Webwork P.S. 1 if you haven't been able yet

Collaborative Exercise

Do the following exercise collaboratively with up to four people per group. Find a population, a sample, the parameter, the statistic, a variable, and data for the following study: You want to determine the average (mean) number of glasses of milk college students drink per day. Suppose yesterday, in your English class, you asked five students how many glasses of milk they drank the day before. The answers were 1, 0, 1, 3, and 4 glasses of milk.

1.2 | Data, Sampling, and Variation in Data and Sampling

Data may come from a population or from a sample. Lowercase letters like x or y generally are used to represent data values. Most data can be put into the following categories:

- Qualitative
- Quantitative

Qualitative data are the result of categorizing or describing attributes of a population. Qualitative data are also often called **categorical data**. Hair color, blood type, ethnic group, the car a person drives, and the street a person lives on are examples of qualitative data. Qualitative data are generally described by words or letters. For instance, hair color might be black, dark brown, light brown, blonde, gray, or red. Blood type might be AB+, O-, or B+. Researchers often prefer to use quantitative data over qualitative data because it lends itself more easily to mathematical analysis. For example, it does not make sense to find an average hair color or blood type.

Quantitative data are always numbers. Quantitative data are the result of **counting** or **measuring** attributes of a population. Amount of money, pulse rate, weight, number of people living in your town, and number of students who take statistics are examples of quantitative data. Quantitative data may be either **discrete** or **continuous**.

All data that are the result of counting are called **quantitative discrete data**. These data take on only certain numerical values. If you count the number of phone calls you receive for each day of the week, you might get values such as zero, one, two, or three.

Data that are not only made up of counting numbers, but that may include fractions, decimals, or irrational numbers, are called **quantitative continuous data**. Continuous data are often the results of measurements like lengths, weights, or times. A list of the lengths in minutes for all the phone calls that you make in a week, with numbers like 2.4, 7.5, or 11.0, would be quantitative continuous data.

Example 1.5 Data Sample of Quantitative Discrete Data

The data are the number of books students carry in their backpacks. You sample five students. Two students carry three books, one student carries four books, one student carries two books, and one student carries one book. The numbers of books (three, four, two, and one) are the quantitative discrete data.

Try It Σ

1.5 The data are the number of machines in a gym. You sample five gyms. One gym has 12 machines, one gym has 15 machines, one gym has ten machines, one gym has 22 machines, and the other gym has 20 machines. What type of data is this?

only makes sense as a whole number
discrete quantitative

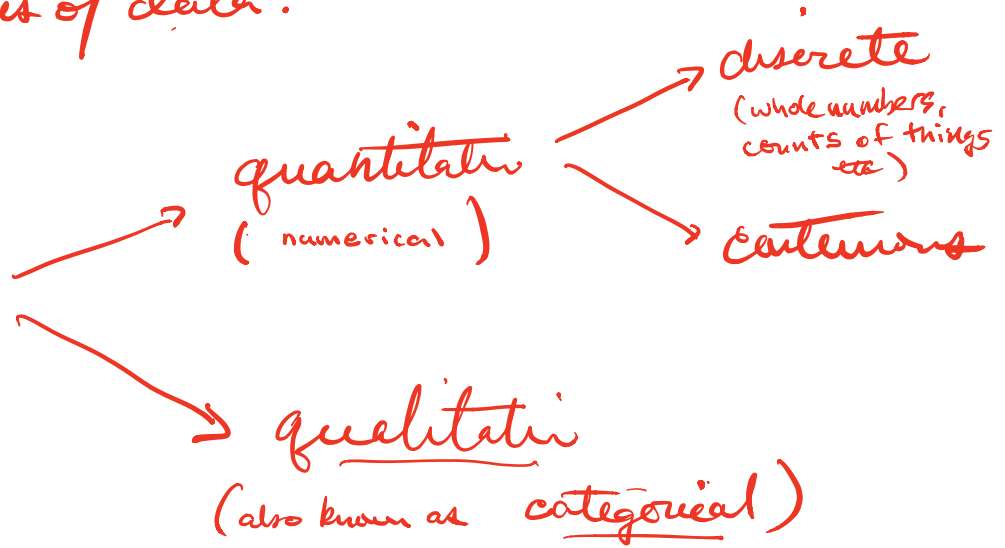
Example 1.6 Data Sample of Quantitative Continuous Data

The data are the weights of backpacks with books in them. You sample the same five students. The weights (in pounds) of their backpacks are 6.2, 7, 6.8, 9.1, 4.3. Notice that backpacks carrying three books can have different

still continuous data
it's possible for weights to have decimals

quantitative continuous data

Types of data:



weights. Weights are quantitative continuous data.

Tip
also, if it has units
it's continuous
(here units are
sq. ft.)

Try It Σ

1.6 The data are the areas of lawns in square feet. You sample five houses. The areas of the lawns are 144 sq. feet, 160 sq. feet, 190 sq. feet, 180 sq. feet, and 210 sq. feet. What type of data is this?

Quantitative Continuous

*- because it could be in decimals
and still make sense*

Example 1.7

You go to the supermarket and purchase three cans of soup (19 ounces tomato bisque, 14.1 ounces lentil, and 19 ounces Italian wedding), two packages of nuts (walnuts and peanuts), four different kinds of vegetable (broccoli, cauliflower, spinach, and carrots), and two desserts (16 ounces pistachio ice cream and 32 ounces chocolate chip cookies).

Name data sets that are quantitative discrete, quantitative continuous, and qualitative.

Solution 1.7

One Possible Solution:

- The three cans of soup, two packages of nuts, four kinds of vegetables and two desserts are quantitative discrete data because you count them.
 - The weights of the soups (19 ounces, 14.1 ounces, 19 ounces) are quantitative continuous data because you measure weights as precisely as possible.
 - Types of soups, nuts, vegetables and desserts are qualitative data because they are categorical.
bisque, lentil, Italian wedding, walnuts, peanuts, broccoli, cauliflower, spinach, carrots, ...
- Try to identify additional data sets in this example.
- *Qualitative data: soup, nuts, vegetables, dessert*

Example 1.8

The data are the colors of backpacks. Again, you sample the same five students. One student has a red backpack, two students have black backpacks, one student has a green backpack, and one student has a gray backpack. The colors red, black, black, green, and gray are qualitative data.

Try It Σ

1.8 The data are the colors of houses. You sample five houses. The colors of the houses are white, yellow, white, red, and white. What type of data is this?

Qualitative (categorical)

NOTE

You may collect data as numbers and report it categorically. For example, the quiz scores for each student are recorded throughout the term. At the end of the term, the quiz scores are reported as A, B, C, D, or F.

Take 5 minutes. 1:09 pm

Example 1.9

Work collaboratively to determine the correct data type (quantitative or qualitative). Indicate whether quantitative data are continuous or discrete. Hint: Data that are discrete often start with the words "the number of."

- the number of pairs of shoes you own
- the type of car you drive
- the distance it is from your home to the nearest grocery store
- the number of classes you take per school year.
- the type of calculator you use
- weights of sumo wrestlers
- number of correct answers on a quiz
- IQ scores (This may cause some discussion.)

quantitative discrete
 qualitative
 quantitative continuous
 quantitative discrete
 qualitative
 quantitative continuous
 quantitative discrete
 quantitative continuous

Solution 1.9

Items a, d, and g are quantitative discrete; items c, f, and h are quantitative continuous; items b and e are qualitative, or categorical.

Try It Σ

1.9 Determine the correct data type (quantitative or qualitative) for the number of cars in a parking lot. Indicate whether quantitative data are continuous or discrete.

Example 1.10

A statistics professor collects information about the classification of her students as freshmen, sophomores, juniors, or seniors. The data she collects are summarized in the pie chart **Figure 1.2**. What type of data does this graph show?

Classification of Statistics Students

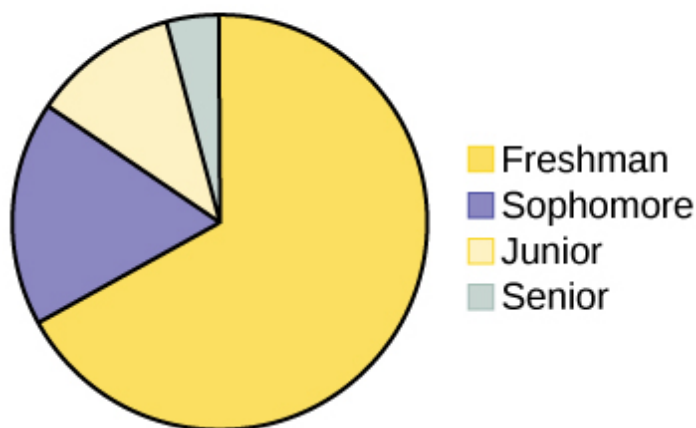


Figure 1.3



Collaborative Exercise

Divide into groups of two, three, or four. Your instructor will give each group one six-sided die. Try this experiment twice. Roll one fair die (six-sided) 20 times. Record the number of ones, twos, threes, fours, fives, and sixes you get in **Table 1.7** and **Table 1.8** (“frequency” is the number of times a particular face of the die occurs):

Face on Die	Frequency
1	
2	
3	
4	
5	
6	

Table 1.7 First Experiment (20 rolls)

Face on Die	Frequency
1	
2	
3	
4	
5	
6	

Table 1.8 Second Experiment (20 rolls)

Did the two experiments have the same results? Probably not. If you did the experiment a third time, do you expect the results to be identical to the first or second experiment? Why or why not?

Which experiment had the correct results? They both did. The job of the statistician is to see through the variability and draw appropriate conclusions.

1.3 | Frequency, Frequency Tables, and Levels of Measurement

Once you have a set of data, you will need to organize it so that you can analyze how frequently each datum occurs in the set. However, when calculating the frequency, you may need to round your answers so that they are as precise as possible.

Answers and Rounding Off

A simple way to round off answers is to carry your final answer one more decimal place than was present in the original data. Round off only the final answer. Do not round off any intermediate results, if possible. If it becomes necessary to round off intermediate results, carry them to at least twice as many decimal places as the final answer. For example, the average of the three quiz scores four, six, and nine is 6.3, rounded off to the nearest tenth, because the data are whole numbers. Most answers will be rounded off in this manner.

I'll try to put a Q in the HW, but wait be an example.

It is not necessary to reduce most fractions in this course. Especially in **Probability Topics**, the chapter on probability, it is more helpful to leave an answer as an unreduced fraction.

Levels of Measurement

The way a set of data is measured is called its **level of measurement**. Correct statistical procedures depend on a researcher being familiar with levels of measurement. Not every statistical operation can be used with every set of data. Data can be classified into four levels of measurement. They are (from lowest to highest level):

- **Nominal scale level**
- **Ordinal scale level**
- **Interval scale level**
- **Ratio scale level**

(always qualitative)

(usually qualitative)

always quantitative continuous

always qualitative (continuous or discrete)

nominal: no natural order to categories

Data that is measured using a **nominal scale** is **qualitative (categorical)**. Categories, colors, names, labels and favorite foods along with yes or no responses are examples of nominal level data. Nominal scale data are not ordered. For example, trying to classify people according to their favorite food does not make any sense. Putting pizza first and sushi second is not meaningful.

Smartphone companies are another example of nominal scale data. The data are the names of the companies that make smartphones, but there is no agreed upon order of these brands, even though people may have personal preferences. Nominal scale data cannot be used in calculations.

Data that is measured using an **ordinal scale** is similar to nominal scale data but there is a big difference. The ordinal scale data can be ordered. An example of ordinal scale data is a list of the top five national parks in the United States. The top five national parks in the United States can be ranked from one to five, but we cannot measure differences between the data.

Another example of using the ordinal scale is a cruise survey where the responses to questions about the cruise are "excellent," "good," "satisfactory," and "unsatisfactory." These responses are ordered from the most desired response to the least desired. But the differences between two pieces of data cannot be measured. Like the nominal scale data, ordinal scale data cannot be used in calculations.

Data that is measured using the **interval scale** is similar to ordinal level data because it has a definite ordering but there is a difference between data. The differences between interval scale data can be measured though the data does not have a starting point.

Temperature scales like Celsius (C) and Fahrenheit (F) are measured by using the interval scale. In both temperature measurements, 40° is equal to 100° minus 60° . Differences make sense. But 0 degrees does not because, in both scales, 0 is not the absolute lowest temperature. Temperatures like -10° F and -15° C exist and are colder than 0.

Interval level data can be used in calculations, but one type of comparison cannot be done. 80° C is not four times as hot as 20° C (nor is 80° F four times as hot as 20° F). There is no meaning to the ratio of 80 to 20 (or four to one).

Data that is measured using the **ratio scale** takes care of the ratio problem and gives you the most information. Ratio scale data is like interval scale data, but it has a 0 point and ratios can be calculated. For example, four multiple choice statistics final exam scores are 80, 68, 20 and 92 (out of a possible 100 points). The exams are machine-graded.

The data can be put in order from lowest to highest: 20, 68, 80, 92.

The differences between the data have meaning. The score 92 is more than the score 68 by 24 points. Ratios can be calculated. The smallest score is 0. So 80 is four times 20. The score of 80 is four times better than the score of 20.

Frequency

Twenty students were asked how many hours they worked per day. Their responses, in hours, are as follows: 5; 6; 3; 3; 2; 4; 7; 5; 2; 3; 5; 6; 5; 4; 4; 3; 5; 2; 5; 3.

Table 1.9 lists the different data values in ascending order and their frequencies.

DATA VALUE	FREQUENCY
2	3

Table 1.9 Frequency Table of Student Work Hours

obvious order to the categories

we know order of categories but no notion of spacing between them

no meaningful distance between categories

For example we cannot say the difference between excellent and good is the same as between satisfactory and unsatisfactory

we can say 100° is 40° hotter than 60°
we cannot say 100° is twice as hot as 50°

100 is twice as many points as 50

1.3 Frequency, Frequency Tables, and Levels of Measurement

39. What type of measure scale is being used? Nominal, ordinal, interval or ratio.

- a. High school soccer players classified by their athletic ability: Superior, Average, Above average **ordinal**
- b. Baking temperatures for various main dishes: 350, 400, 325, 250, 300 **interval**
- c. The colors of crayons in a 24-crayon box **nominal**
- d. ~~Social security numbers~~ **nominal** (sorting by social security number is not measuring anything; for all intents and purposes, SSN is a label, like a name)
- e. Incomes measured in dollars **Ratio**
- f. A satisfaction survey of a social website by number: 1 = very satisfied, 2 = somewhat satisfied, 3 = not satisfied **Ordinal**
- g. ~~Political outlook: extreme left, left of center, right of center, extreme right~~ **Nominal**
- h. Time of day on an analog watch **Interval**
- i. The distance in miles to the closest grocery store **Ratio**
- j. The dates 1066, 1492, 1644, 1947, and 1944 **Interval**
- k. The heights of 21–65 year-old women **Ratio**
- l. Common letter grades: A, B, C, D, and F **Ordinal**

e. Ratio because it makes sense to say "You made twice as much as me"

i. Ratio because it makes sense to say "That distance is twice this distance"
"This grocery store is twice as far as that grocery store"

j. 1500 is 500 years after 1000 ✓
"1500 is 50% later than 1000" doesn't make sense

l. Ordinal because "A is higher than B" ✓
but we can't say "A is higher than B
by the same amount that
D is higher than F"